



PROJECT REPORT
DATA-DRIVEN ATHLETE PERFORMANCE
GROUPING: A COMPARATIVE STUDY OF K-MEANS AND
K-MEDOIDS WITH MKNN ACCURACY ASSESSMENT

ADONIS MAX BALINDA
22.K1.0004

Faculty of Computer Science
Soegijapranata Catholic University
2026

ABSTRACT (ABSTRACT TITLE)

This study proposes a hybrid machine learning approach to classify athlete fitness levels from unlabeled physical activity data by integrating clustering and classification techniques. The research addresses the absence of predefined fitness labels in wearable-based fitness datasets. The dataset used is the Workout & Fitness Tracker Dataset consisting of 10,000 records with several numerical attributes such as Steps Taken, Calories Burned, Distance, Workout Duration, and Daily Calories Intake.

In the first stage, K-Means and K-Medoids clustering algorithms were applied to group the data and generate pseudo-labels representing different fitness levels. The clustering results were evaluated using the Silhouette Score to measure cluster quality. In the second stage, the generated pseudo-labels were used to train a Modified K-Nearest Neighbor (MKNN) classification model.

The experimental results show that clustering-based pseudo-labeling can effectively support classification on unlabeled fitness datasets. The proposed hybrid approach demonstrates promising performance in grouping athlete fitness levels and provides a data-driven method for analyzing physical activity data.

Keywords: Athlete Fitness, Clustering, MKNN, K-Means, K-Medoids, Machine Learning

